

WP3 – Information Platform

Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências, Departamento
de Informática

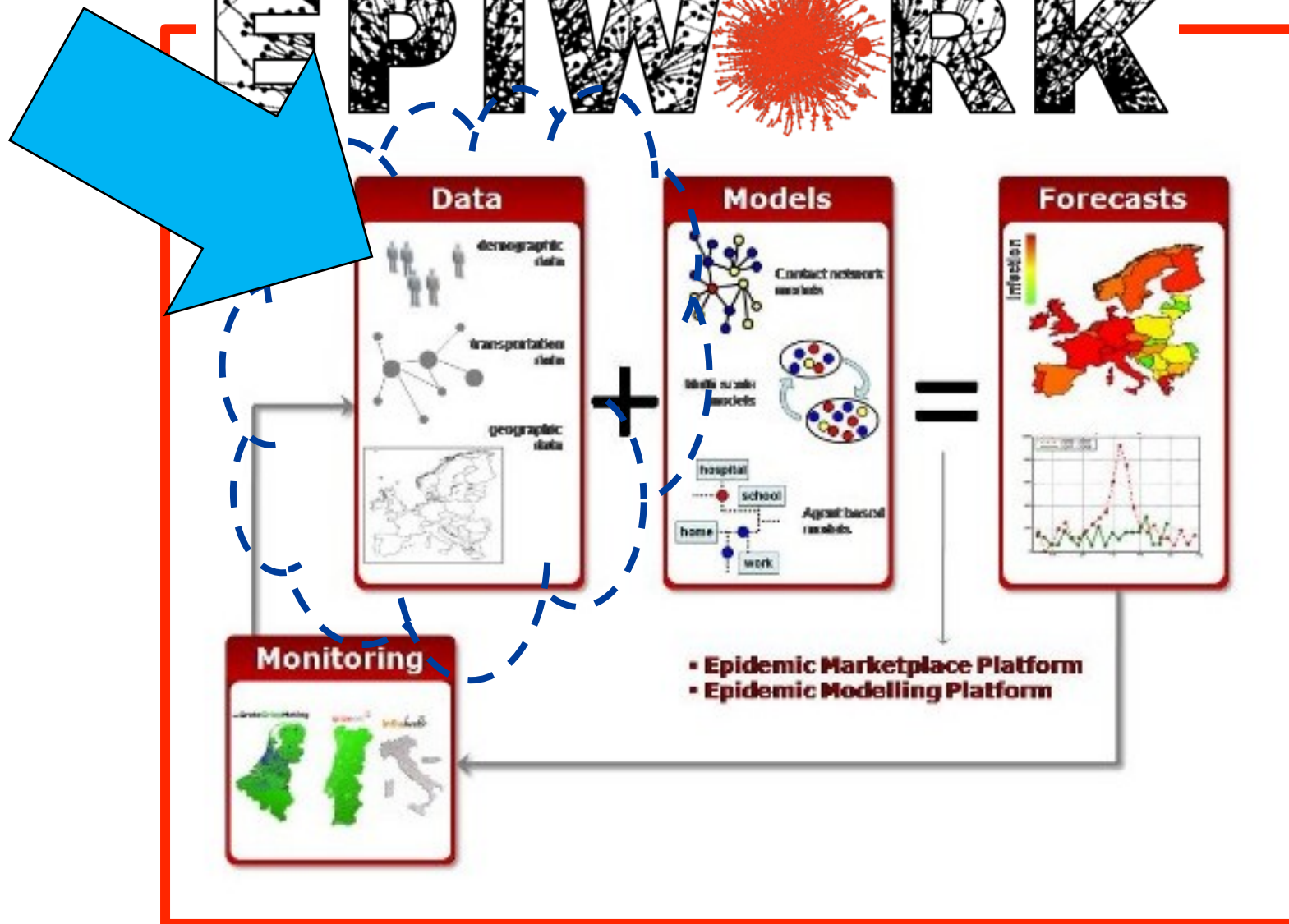
mjs@di.fc.ul.pt



What will be necessary to predict epidemics precisely?

- **Data** of many different types and many unrelated sources.
 - Improved accuracy makes required data a never-ending story
 - We all want to see realistic and timely plots of epidemics propagation.
 - Available, but **hard to find, collect and maintain!**





EPIWORK

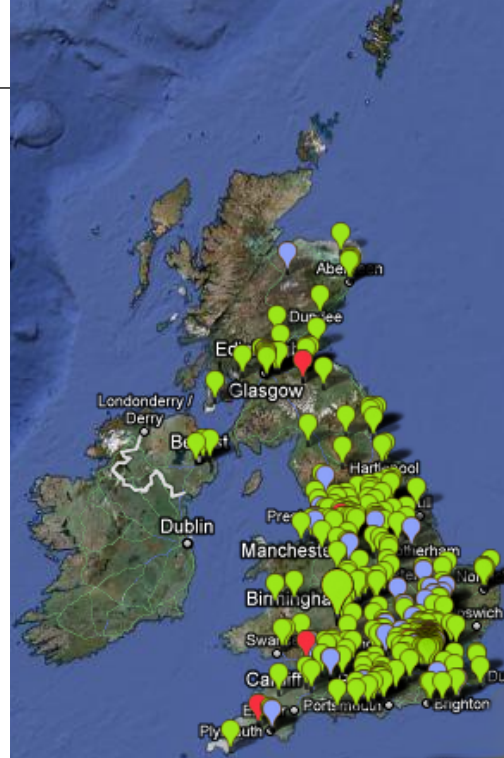


Número de participantes

Portugal		5477 participantes
Holanda		20906 participantes
Bélgica		5839 participantes
Itália		3019 participantes
México		3950 participantes
Brasil		370 participantes
Reino Unido		5551 participantes
Australia		7776 participantes
Montreal (CA)		951 participantes

deGroteGriepMeting.nl

-  meting (15444)
-  verkouden (3602)
-  vermoedelijk griep (87)
-  grens postcode gebied



gripenet 

<http://www.gripenet.pt/>

INFLUWEB 

 persone senza sintomi  persone con raffreddore  persone con sintom



Other Internet Monitoring Sources

google.org Flu Trends

[Google.org home](#)

Flu Trends

Japan

[Download data](#)

[Home](#)

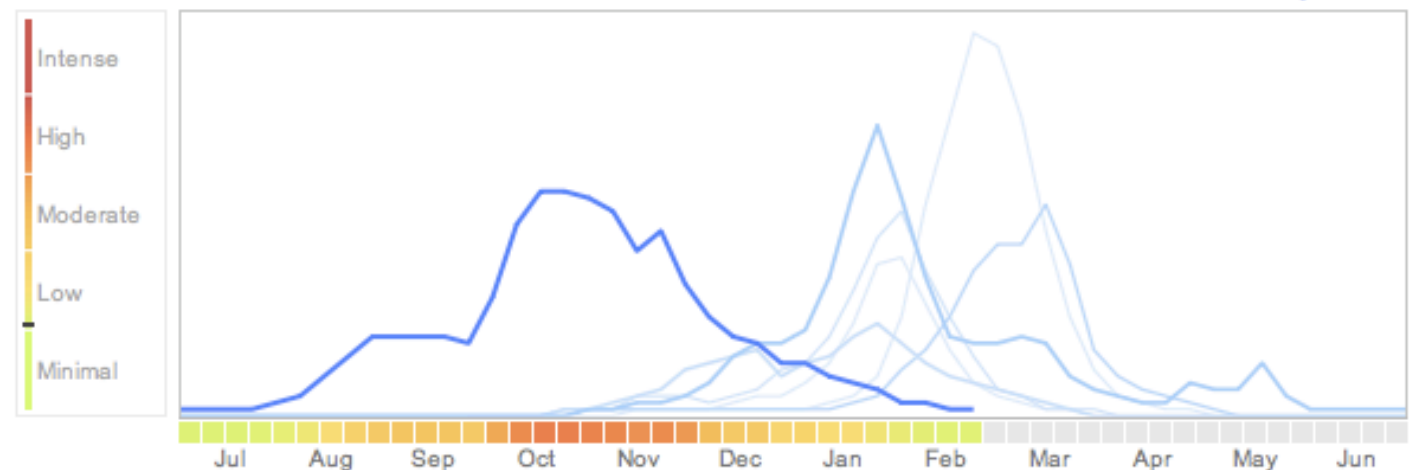
[How does this work?](#)

[FAQ](#)

Explore flu trends - Japan

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

National



Social Media Sources



disease	location	source	author	evidence	score	date
HIV	Italy	Twitter	HumanityNews (Humanitarian News)	News: Patent pool decision heralds era of cheap HIV drugs http://dlvr.it/L5Bq	0	2010-03-22 10:48:14
HIV	Italy	Twitter	USArmyAfrica (USArmyAfrica)	News Update Africa: Steps in the right direction against HIV/ AIDS : Africa's battle against HIV/ AIDS is not ... http://bit.ly/acEbGC #Africa	0	2010-03-22 09:19:32
HIV	Italy	Twitter	romavisibile (Roma Visibile)	Social Romavisibile L' Aids lavorando per il guru e la setta mi ha abbandonata http://ow.ly/16Rhcl	0	2010-03-22 09:16:44
HIV	Italy	Twitter	Grantun19 (Grant Vaughn)	Safety concern for HIV drug combination http://is.gd/aSBEL	0	2010-03-22 08:37:05
HIV	Italy	Twitter	NonProfitBlogs (NonProfitBlogs)	Aids warning over bushmeat trade http://dlvr.it/L31N	0	2010-03-22 06:54:17
HIV	Italy	Twitter	HumanityNews (Humanitarian News)	Blogpost: Aids warning over bushmeat trade http://dlvr.it/L2Gb	0	2010-03-22 05:22:13
HIV	Italy	Twitter	casso89 (Antonino Cassotta)	Farmaco contro l'acne previene riattivazione di HIV latente - http://wp.me/pObs6-6l	0	2010-03-21 22:01:25
HIV	Italy	Twitter	mwbloem (Martin Bloem)	GWU and WFP organized a meeting on nutrition/food security in HIV/ AIDS programs; excellent input to the new WFP's policy on HIV AIDS .	0	2010-03-21 19:08:38
HIV	Italy	Twitter	duduramone (EDUARDO N. VANTINI)	@niderodriguez tomara que vc morra corroido pela aids	0	2010-03-21 17:45:30
HIV	Italy	Twitter	notiziegay (Notizie gay)	New post: Cinema. Perché il rischio Aids non va riposto nel cassetto. Un film-documentario itali (http://cli.gs/mdYEG) http://cli.gs/mdYEG	0	2010-03-21 09:42:16

data.gov.uk^{BETA}

Opening up government

What are you looking for?

Search



Data

Apps

Ideas

Forum

Wiki

Blogs

Linked Data

Resources

About

Search

Enter your keywords:

epidemiology

Search

Did you mean:

audiology

Search results



Quarterly Epidemiological Commentary

Dataset

UK Statistics Authority **Epidemiological** analyses of Mandatory surveillance data on MRSA bacteraemia and C. difficile infection Source agency: Health Protection Agency Designation: Official Statistics not designated as National Statistics ...

► [UK Statistics Authority \(UKSA\)](#)



Quarterly Epidemiological Commentary

Dataset

Health Protection Agency **Epidemiological** analyses of Mandatory surveillance data on MRSA bacteraemia and C. difficile infection Source agency: Health Protection Agency Designation: Official Statistics not designated as National Statistics ...

Current search

Search found 11 items

» (-) epidemiology

Sort by

» Relevancy

» Title

» Type

» Last updated

» Rating

Filter by type

» Dataset (11)



[HOME](#) [DATA](#) [TOOLS](#) [COMMUNITY](#) [METRICS](#) [DIALOGUE](#)

epidemiology

Search

[Advanced Search](#)

Results 1-10 of about 10 for 'epidemiology'

Web results by

Related Topics

- [Epidemiology of Cancer](#)
- [Health](#)
- [Medicine](#)
- [Surveillance Epidemiology and End Results](#)

Cancer Incidence - Surveillance, **Epidemiology**, and End Results ...

SEER Limited-Use cancer incidence data with associated population data. Geographic areas available are county and SEER registry. The Surveillance, **Epidemiology**, and ...

<http://www.data.gov/.../119> - Cached

Cancer Mortality in the United States: 1970-1994 - Direct Download

The data are from a cancer mapping project in theNational Cancer Institute's (NCI) Division of Cancer **Epidemiology** andGenetics (DCEG).Purpose:These data are intended for ...

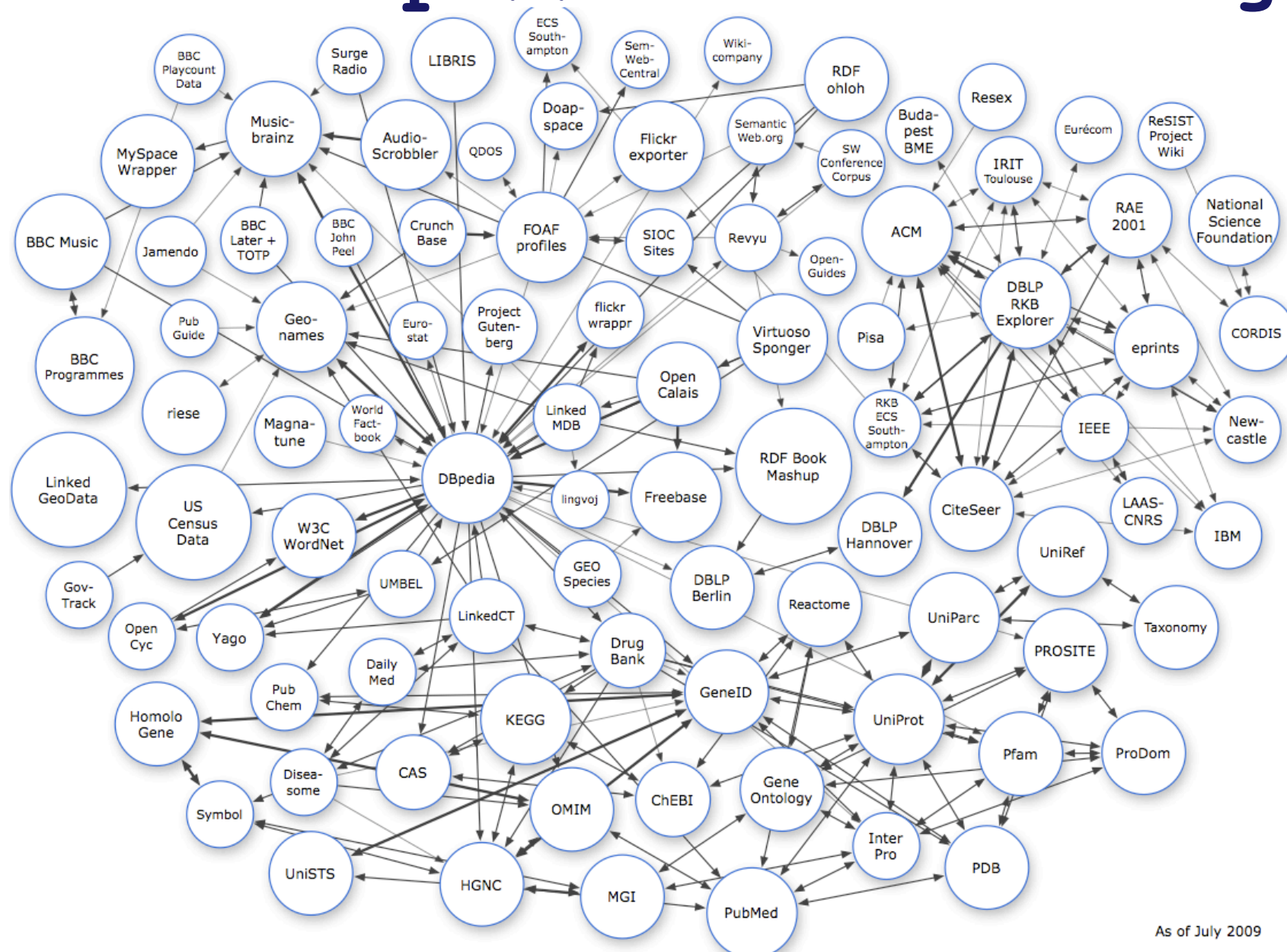
<http://www.data.gov/.../g602094> - Cached

Air Quality System (AQS) ambient observations: 2007 ozone

... Quality, Air Pollution, EPA, ambient air, public health, environment, federal data download, clean air trends, monitoring air, federal datasets, respiratory, **epidemiology**

<http://www.data.gov/.../1127> - Cached

<http://linkeddata.org/>



Data in EPIWORK

Classic Sources

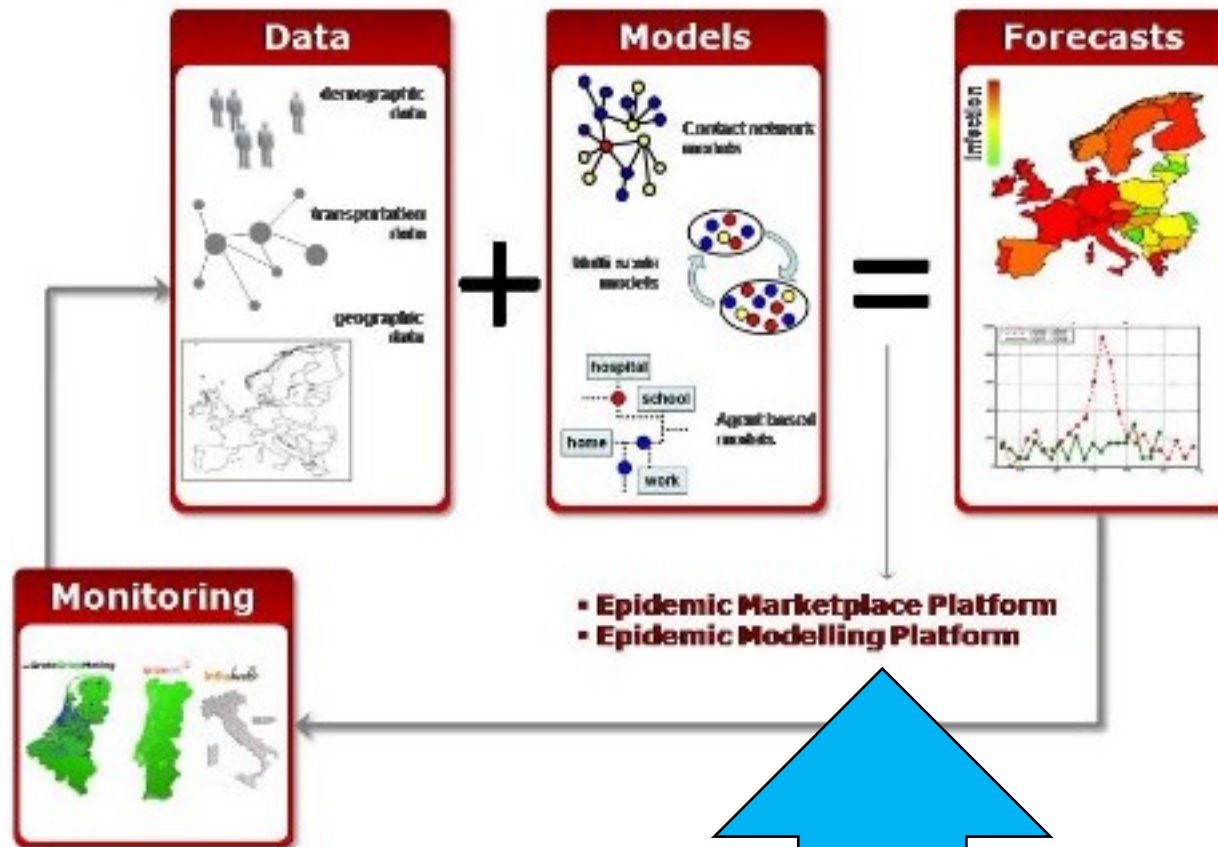
- [National Bureau of Statistics] demographics, transportation data, ..
- [Public Health authorities] surveillance data (maybe?)

Modern Sources

- [Internet Monitoring Sources]
- [Social Media] behavioural data

To be shared by epidemic modellers in a digital library, dubbed the **Epidemic Marketplace**

EPIWORK



- Epidemic Marketplace Platform
- Epidemic Modelling Platform



Outline

1. The need for an Epidemic Marketplace
2. **Epidemic Marketplace 1.0**
3. D3.3 Public Release of the Epidemic Marketplace Platform
4. Where we stand and plans for work ahead

Steps for Creating the EM

1. Elaborate **meta-model** for describing datasets used by epidemic modellers.
2. Provide **query services** over the meta-data to discover resources.
3. Select **ontologies** for characterizing data and develop an ontology of epidemic concepts.
4. Ingest, harmonize and **cross-link data**.
5. Provide **query services to select epidemic data** using the EM meta-data and ontologies.

Common Reference Model

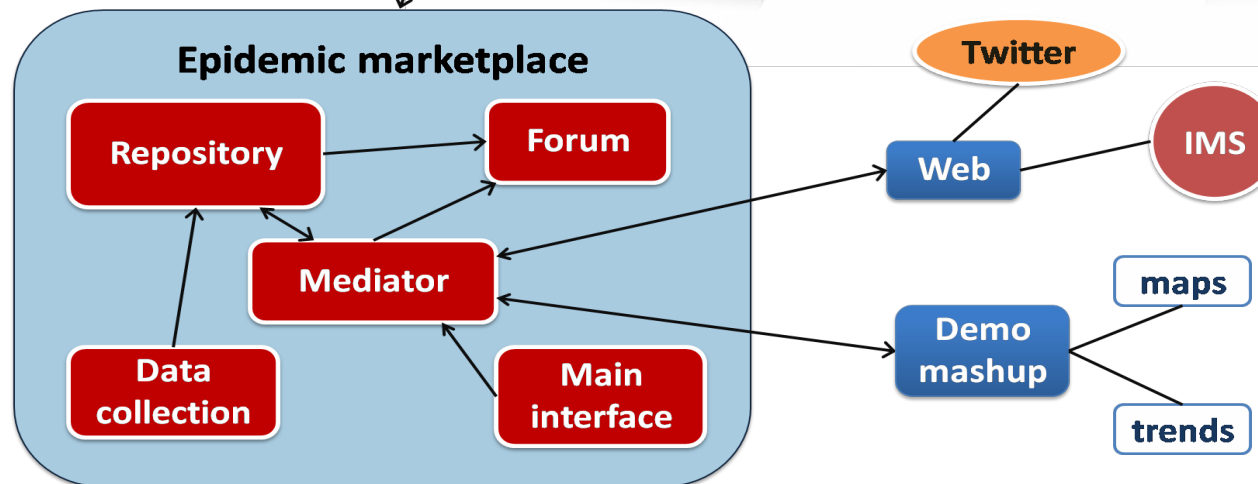
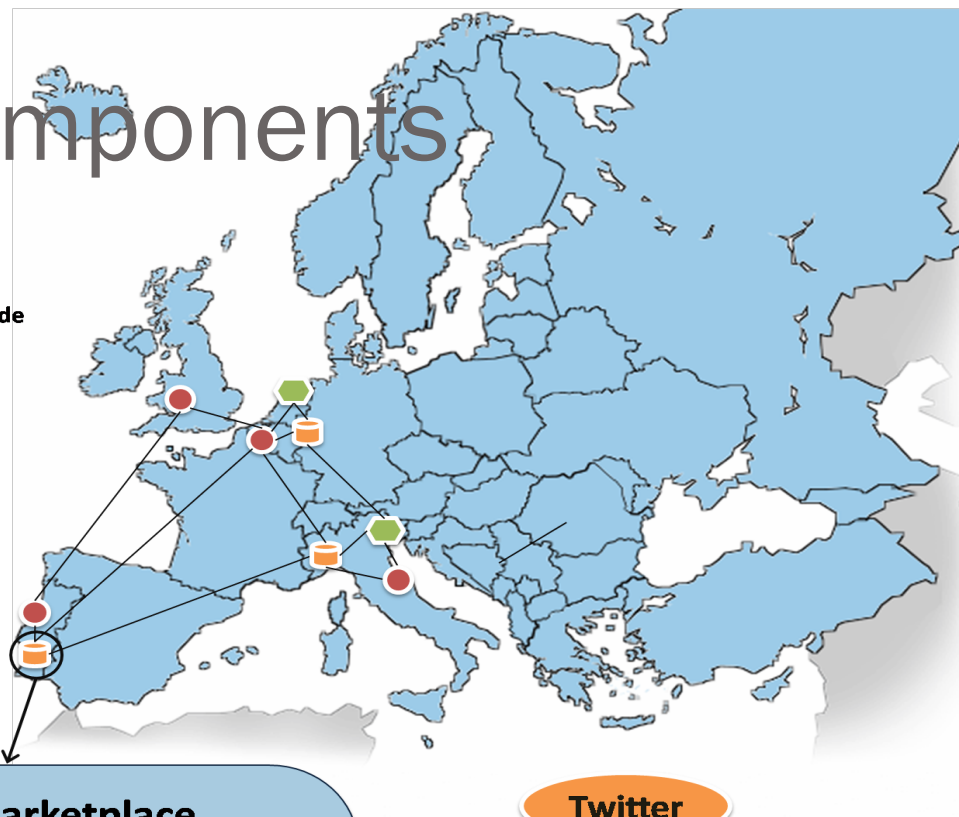
- **Open domain:** detailed description of the datasets used in the models of all sorts of epidemics would **require describing virtually every kind of information**, given the diversity of factors and the interdisciplinary of epidemiologic studies.

Data model needs to support **interlinked data**.

Meta-data and Ontologies

- The **information model** of the EM is directly **defined as metadata and ontologies**.
- Ontology and Meta-data standards, the Pros and Cons of using them, annotation and deployment strategies, and the steps for creating an metamodel for epidemic data were the subject of D3.1 reviewed last year.

EM: Main Components



EM 1.0 Software Components

- Fedora Commons 2.X for the implementation of the main features of the repository.
- Access control in the platform
 - XACML (OASIS 2010),
 - LDAP (Tuttle et al. 2004)
 - Shibboleth (identity management).
- Front-end based in Muradora
- Forum based on phpBB (+ Muradora)



Outline

1. The need for an Epidemic Marketplace
2. Epidemic Marketplace 1.0
3. **D3.3 Public Release of the Epidemic Marketplace Platform**
4. Where we stand and plans for work ahead

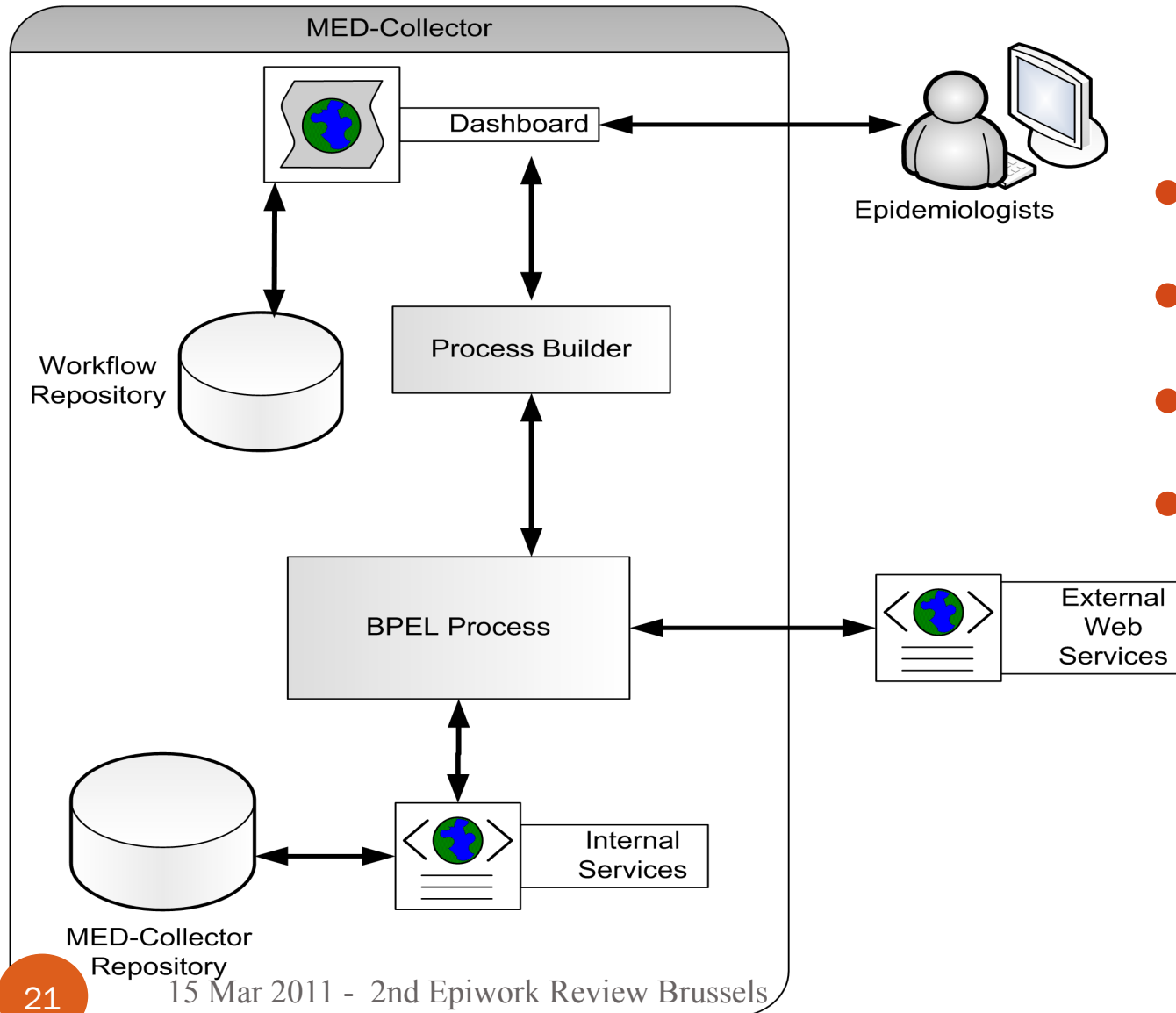
What is new since Mar 2010?

1. Improved reliability
2. MEDCollector – automatic data collector
3. Meta-data policies and editor
4. Web services API + Simple EM Client
5. Improved user interface
6. Public: anyone can browse and register
(required for upload)

Improved Reliability

- Reorganizations and back-end Services Before Public Deployment
 - Virtualized environment: every major component running on two separate virtual machines - production + development environments (Xen+CentOS)
 - Monitoring and alerts for all services (Nagios)
- Logging and Analysis (Google Analytics)

MEDCollector

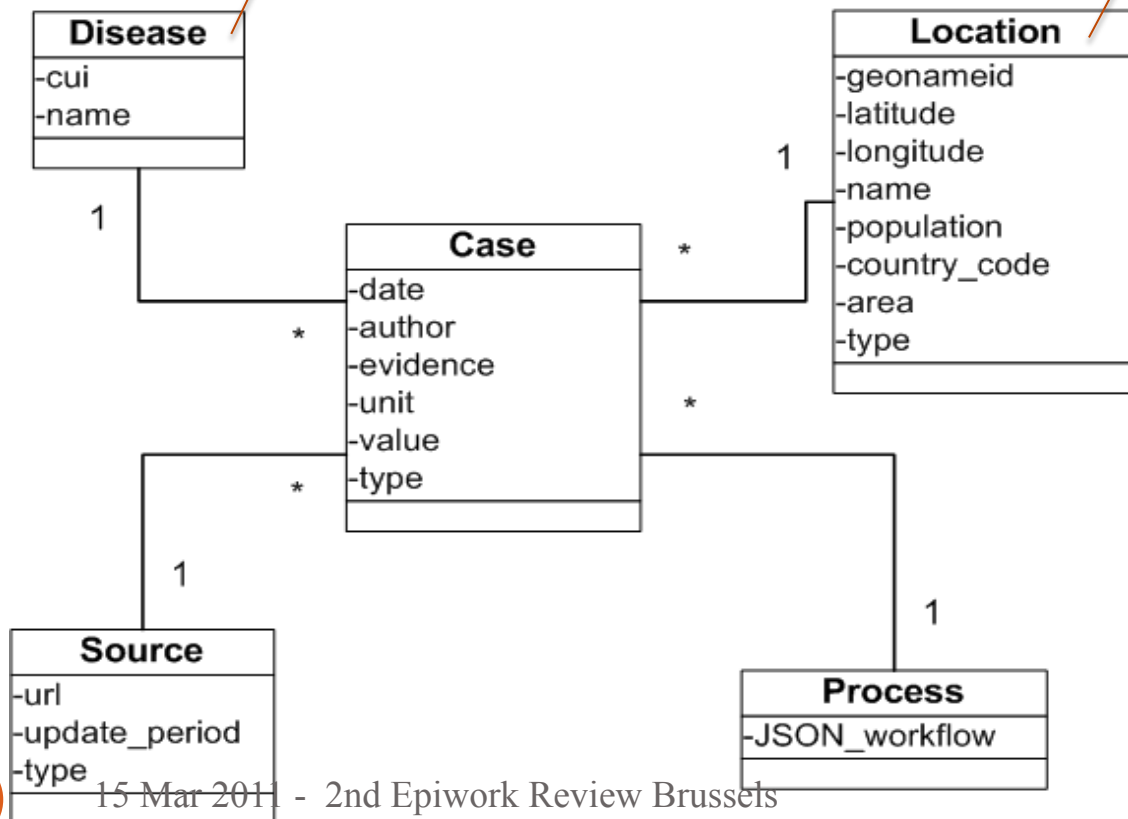


- Web Services
- Workflow Processes
- Local Storage
- Dashboard for Workflow Design

MEDCollector Data Model

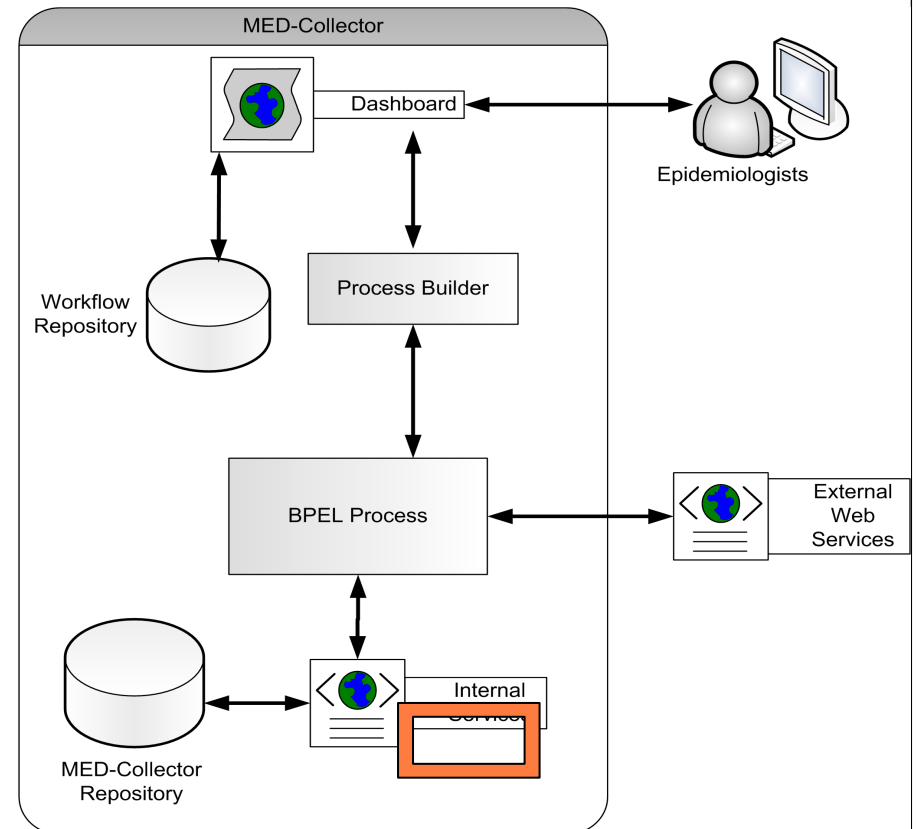
UMLS: twitter searched ion
subset of 89 terms
related to "Disease or
Syndrome"

Geonames.org:
All Countries
and Capitals



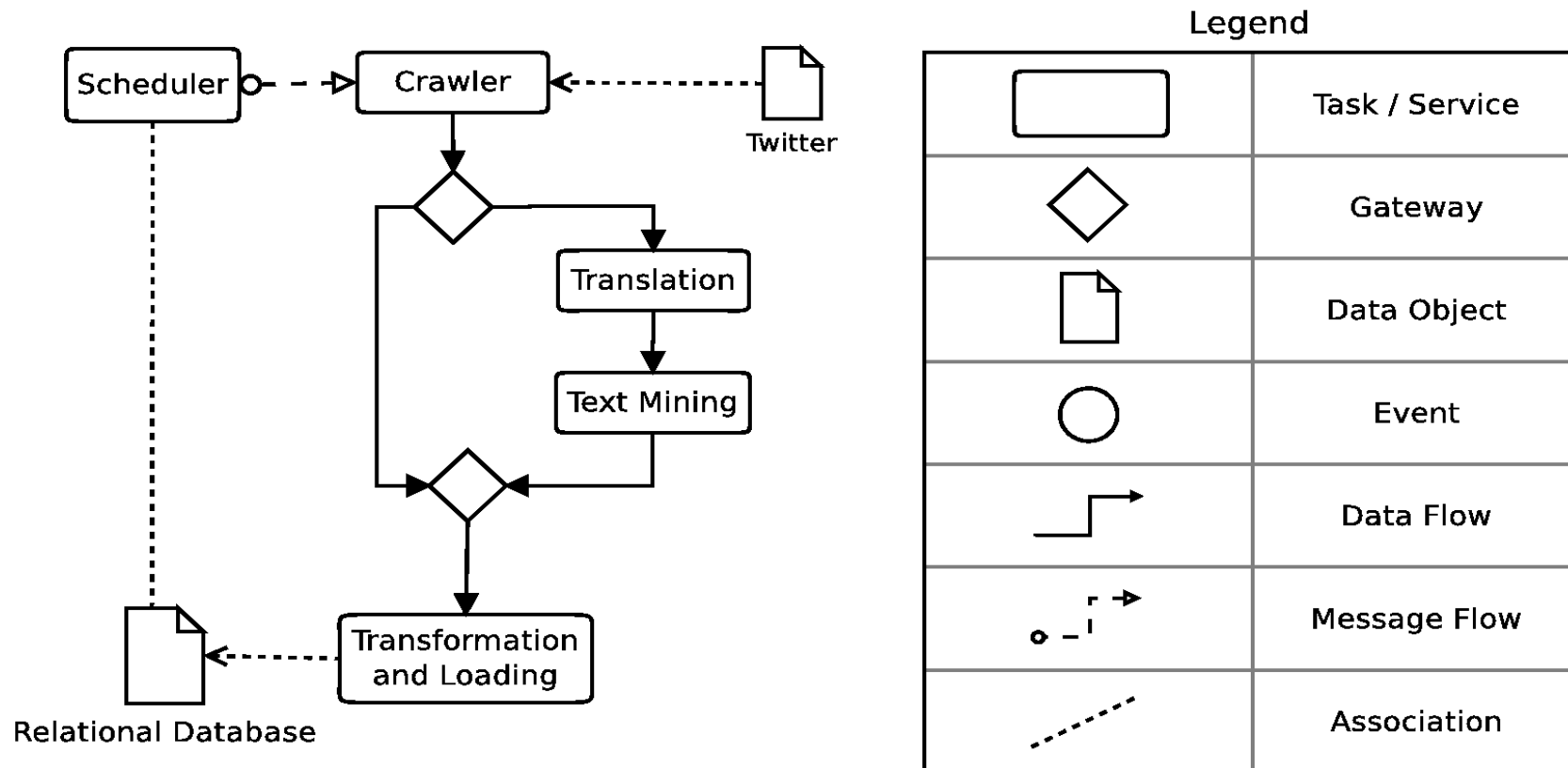
MEDCollector Services

- Data Collection Services
 - Query Selection Services
 - Data Harvesting Services
 - XML Transformation Services
 - Database Loading Service
- Data Packaging Services
 - To CSV



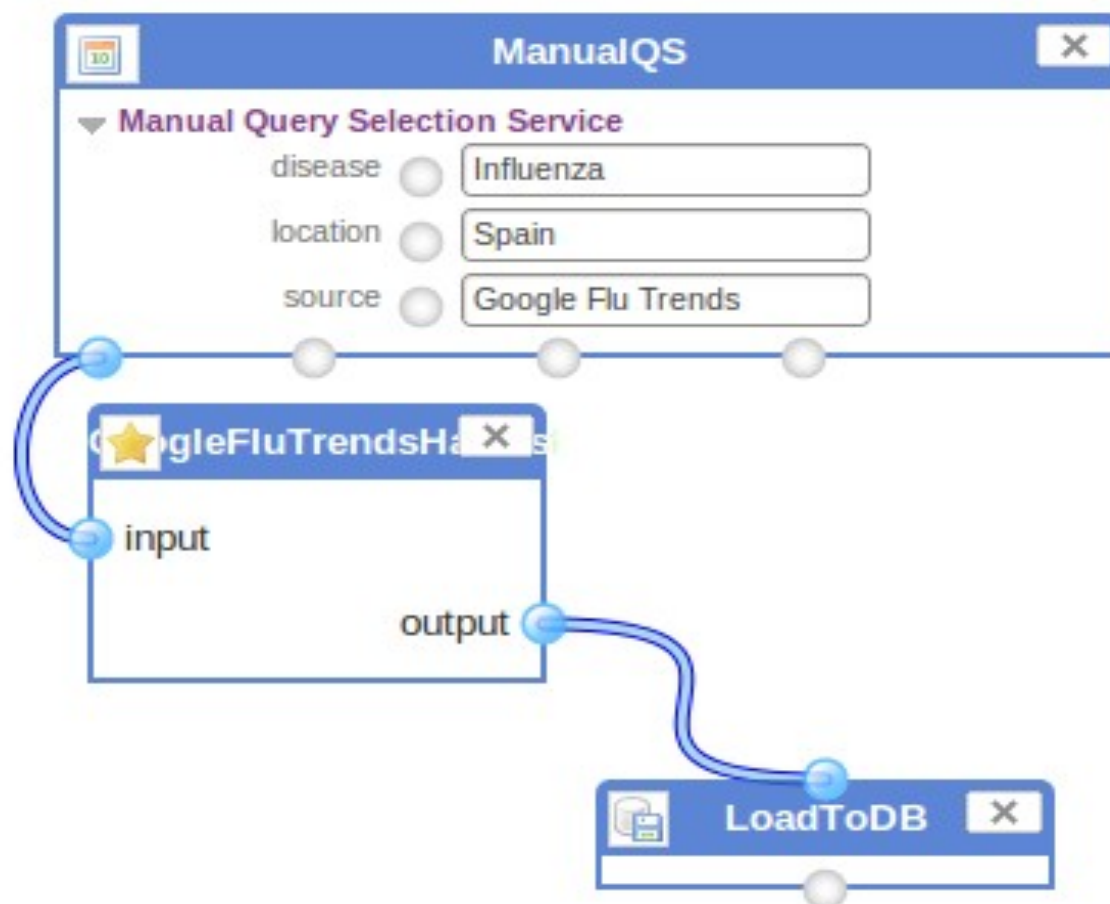
MEDCollector - BPEL

Language to define how Web-Services Communicate



Standard graphical notation – BPMN → Complex!

MEDCollector: Dashboard



MEDCollector: Dashboard

The screenshot displays the MEDCollector Dashboard, titled "Epidemic Marketplace Data Collection". The interface includes a top navigation bar with tabs for "Data Collection", "Data Packaging", and "View Workflows". A welcome message for "João Zamite" is visible on the right, along with links for "logout" and "Epidemic Marketplace".

The main section is titled "Data Collection" and features a toolbar with buttons for "New", "Load", "Save", "Run", "Delete", "Help", and "Feedback". On the left, a "Modules" sidebar lists options: "Query Selection", "Crawlers", "XML", "Text", "Loading", "Date", and "Other".

The central workspace shows two configuration windows:

- RESTfulWebService**: Contains a "URL" field with the value "http://www2c.cdc.gov/podcasts/createrss.asp?t=r&c=20" and several empty fields for "Query", "Parameter Name", and "Parameter Value".
- RSSToRecord**: A sub-window with a "Disease and Location are optional" section. It includes fields for "Disease" (Influenza), "Location" (United States), and "Source" (CDC Flu updates).

On the right, a "Properties" panel shows configuration for "CDCFluRSS", including a "Description" field, a "Scheduling" dropdown set to "Automatic", and a "period (minutes)" field set to "720". Below this is a "Minimap" section and a "View/Add Entities" table with buttons for "View Diseases", "Add Disease", "View Sources", and "Add Source".

Watch the Demo!

Automatically Collected Data

- Twitter: 89 diseases, world-coverage
- ProMed-mail
- Google Flu Trends
- CDC RSS Feeds
 - Flu updates
 - Travel Notices
 - ...
- Periodically packed and uploaded to the EM repository

What is new since Mar 2010?

1. Improved reliability
2. MEDCollector – automatic data collector
3. Meta-data policies and editor
4. **Web services API + Simple EM Client**
5. Improved user interface
6. Public:
anyone can browse and register (for upload)

Meta-data Policies and Editor

- Meta-data introduction **simplified**
 - Editor that pops-up on upload now fills most of the entries with appropriate defaults.
- EM Repository Meta-data Vocabulary
 - Generic DCTERMS **adopted** for datasets characterisation
 - Epidemics-specific DCTERMS **defined** for epidemic datasets characterisation

The screenshot displays a web-based form for entering meta-data. It features several sections, each with a plus icon and a question mark for expansion. The 'Title' section has a red 'x' icon and a text input field containing 'US Data.Gov'. The 'Subject' section has a red 'x' icon and four dropdown menus with the following selected values: 'Demographic', 'Epidemic', 'Economic', and 'Environment'. The 'General Description' section is currently collapsed. The 'Organisation' and 'Publisher' sections are also collapsed. The 'Spatial coverage' section has a red 'x' icon and three input fields: 'Country' (containing 'United States'), 'City' (empty), and 'Region' (empty).

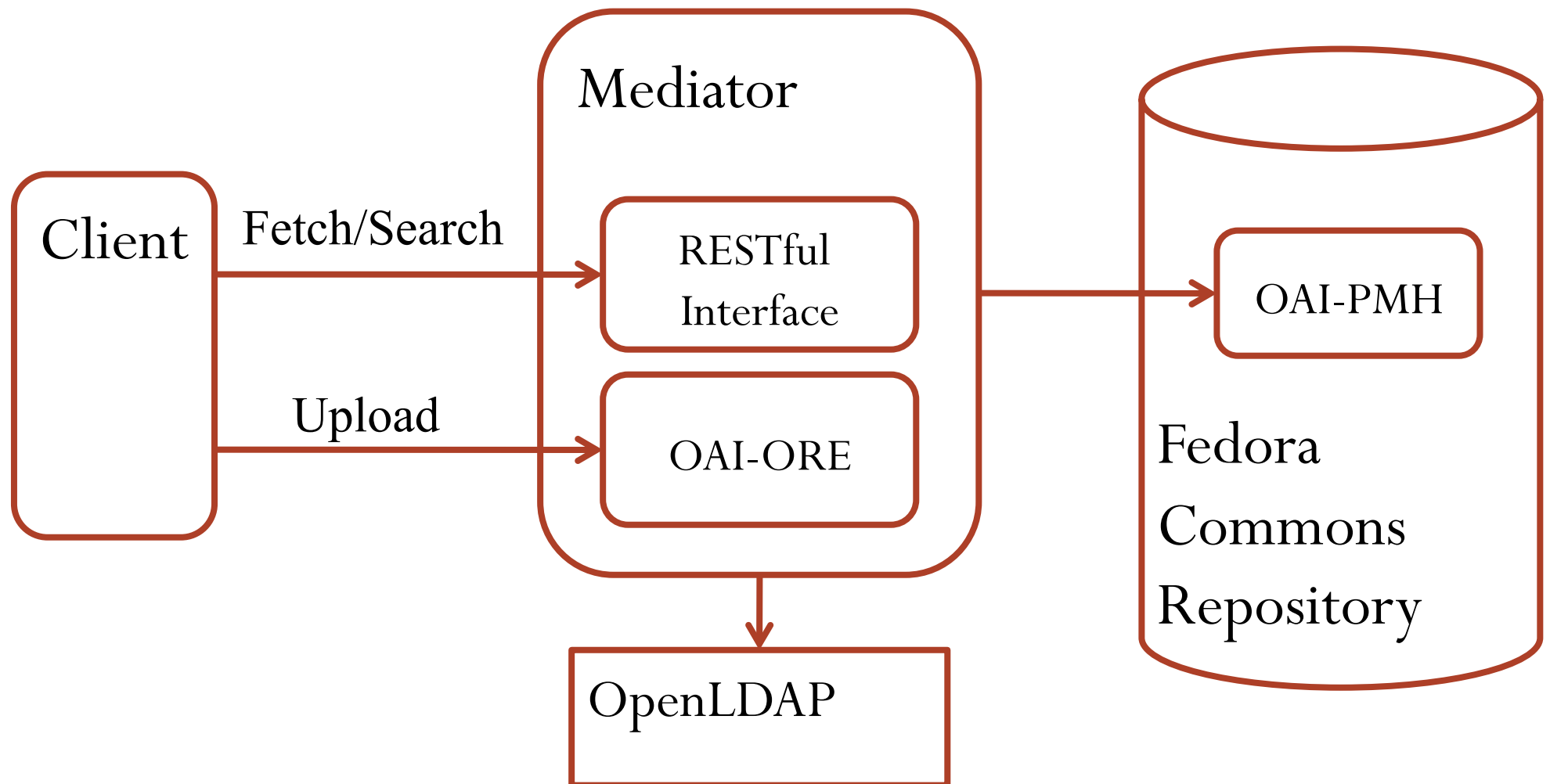
DC Term Example: RightsHolder

Name of Term	RightsHolder
EM Term URI	http://epimarketplace.net/metadata/em-terms/rightsHolder
EM qualified name	em:rightsHolder
DC Term URI	http://purl.org/dc/terms/rightsHolder
DC Qualified name	dc:rightsHolder
Label	Rights Holder
Defined By	http://dublincore.org/documents/dcmi-terms
DC Definition	A person or organization owning or managing rights over the resource.
EM Definition	
DC Comments	
EM Comments	
Refines	http://epimarketplace.net/metadata/em-terms/rights
Refined By	
Has Encoding Scheme	
Obligation	RA
Occurrence	0 or more

EM Term Example: HostGroup

Name of Term	HostGroup
EM Term URI	http://epimarketplace.net/metadata/em-terms/hostGroup
EM qualified name	em:hostGroup
Label	Host Group
Defined By	http://epimarketplace.net/metadata/em-terms
EM Definition	A dataset may contain data not about a specific species but rather a larger group comprising different species. When the larger group is the target this element should be used. For example, when referring to groups such as rodents or household animals.
EM Comments	When possible a scientific classification should be used.
Refines	http://epimarketplace.net/metadata/em-terms/epidemiological
Refined By	
Has Encoding Scheme	
Obligation	RA
Occurrence	0 or more

Mediator Web Services



Simple EM Client

Download from <http://epimarketplace.net/mediator/>

```
Welcome to the Simple EM Client.
```

```
Username: mjs
```

```
Password:
```

```
Login successfull. Welcome mjs
```

```
Menu:
```

```
g. Get Resource
```

```
l. Show Resource List
```

```
L. Show Collection List
```

```
a. Add Resource
```

```
A. Add Collection
```

```
r. Remove Resource
```

```
R. Remove Collection
```

```
d. Download Updates
```

```
u. Upload Resource
```

```
Insert an option ("e" or "q" to leave):
```

- Mapping of client filenames to EM resources (FC data streams and Collections)
- Operations:
Check-out, check-in

Watch the Demo!


[Home](#)

Resource Metadata

Meta-data	
Title	Twitter dataset H1N1 + Portugal 4-6-2009
Creator	LASIGE node of the Epidemic Marketplace
Subject	twitter message dataset
Description	<p>This dataset contains Twitter messages containing the words H1N1 and Portugal collected between 16-5-2009 and 3-6-2009, Information is a 7 columns relation, containing the following data:</p> <p>Column 1- keyword 1 (disease)- H1N1 Column 2- Keyword 2 (location)- Portugal Column 3- Source (Twitter) Column 4- Author of the message (user id) Column 5- The message body (evidence) Column 6- score Column 7- date (day and hour)</p>
Publisher	Epiwork – http://www.epiwork.eu
Format	text/tab-separated-values
Language	English, Portuguese
Contributor	Luis F Lopes, Joao M Zamite, Bruno C Tavares, Francisco M Couto, Fabricio Silva, Mario J Silva
Relation	Luis F. Lopes, João M. Zamite, Bruno C. Tavares, Francisco M. Couto, Fabrício Silva and Mário J. Silva. (2009). Automated Social Network Epidemic Data Collector. INForum informatics symposium.
Source	http://epiwork.di.fc.ul.pt/collector/
Coverage	Spatial: Portugal, Temporal: 2009-5-16 to 2009-6-3
Rights	Creative Commons Attribution-ShareAlike (CC BY-SA), http://creativecommons.org/licenses/by-sa/3.0/

Old Graphic Style

Epidemic Marketplace

// Epidemic Marketplace

// Datasets

// Forum

// Web Services

// Data Collection

// Contact

// Learn More

**Epidemic
Marketplace**

... a platform for integrating and sharing epidemiological data.

Try it at:
<http://epimarketplace.net>

Some features of the Epidemic Marketplace require user authentication. Please [register](#)

Datasets

Forum

Web Services

Data Collection

Announcements

2010.10.01 // Epimarketplace.net is now open to user registrations from the entire community. [\(See More\)](#)

Outline

1. The need for an Epidemic Marketplace
2. The Epidemic Marketplace
3. D3.3 Public Release of the Epidemic Marketplace Platform
4. **Where we stand and plans for work ahead**

WP3: status (what we have done)

- Deliverable D3.1 (meta-model) released
- Deliverable D3.2 (prototype) released
 - Hardware and base software deployed;
 - Initial prototype of EM with initial set of characterized datasets
- Deliverable D3.3 (public version) released
 - Data-collector
 - EM DCAP and meta-data handling
 - Web Services

Events 2nd year

London, Delhi, Bilbao, ERCIM News



Privacy in Socially Intelligent ICT

Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências,
Departamento de Informática

mjs@di.fc.ul.pt

Towards a Science of Socially Intelligent ICT
Imperial College London, 3rd August 2010



ECCS'10 LISBON
European Conference on Complex Systems
eccs2010.eu



EM-related Publications (2nd year)

1. Mário J. Silva, Fabrício A.B. Silva, Luís Filipe Lopes, Francisco M Couto, **Building a Digital Library for Epidemic Modelling**. Proceedings of ICDL 2010 - The International Conference on Digital Libraries 1, p. 447–459, New Delhi, India, 23–27 February, 2010. TERI Press–New Delhi, India. Invited Paper.
2. Fabrício A.B. Silva, Mário J. Silva, Francisco M Couto, **Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis**. ERCIM News 82 – Special Theme: Computational Biology. July, 2010.
3. Luis Filipe Lopes, Fabrício A.B. Silva, Francisco M Couto, João Zamite, Hugo Ferreira, Carla Sousa, Mário J. Silva, **Epidemic Marketplace: An Information Management System for Epidemiological Data**. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.
4. João Zamite, Fabrício A.B. Silva, Francisco M Couto, Mário J. Silva, **MEDCollector: Multisource Epidemic Data Collector**. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.
5. João Zamite, **Multisource Epidemic Data Collector, Master Dissertation**, University of Lisbon, Faculty of Sciences, September 2010.
6. Luis Filipe Lopes, **A Metadata Model for the Annotation of Epidemiological Data**, Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.
7. Hugo Ferreira, **O Mediador do Epidemic Marketplace**. Master Dissertation, University of Lisbon, Faculty of Sciences, September, 2010; (in Portuguese).

WP3: status (what we will do)

- Overcoming the initial difficulties in hiring the planned resources
 - Refreshed team with competencies required for the 2nd and 3rd year; Hiring 1 sw eng for push in release of EM 2.0
- Working on Epidemic Marketplace 2.0
- D3.4 and D3.5 due Feb 2012
 - site analytics
 - interlinking
- Peeking on how to address challenges for the 4th year
 - negotiating access to content

Changes in UL WP3 Team

Out

- Fabrício Silva
- Luis F. Lopes (meta-data)
- Hugo Ferreira (mediator)

In

- Dulce Domingos (access control)
- Juliana Duque (information architecture, graphics)
- João Ferreira (ontologies)
- + (always in)
- Mário
- Francisco
- João Zamite

Scheduled Deliverables

		Year 1			Year 2			Year 3			Year 4		
		M4	M8	M12	M16	M20	M24	M28	M32	M36	M40	M44	M48
WP3	Information platform												
Task 1			D3.1										
Task 2			D3.1										
Task 3				D3.2		D3.3				D3.4			D3.6
Task 4										D3.4 D3.5			D3.6

15 Mar 2011

Todo List and Planning (Brussels, Mar 2011)

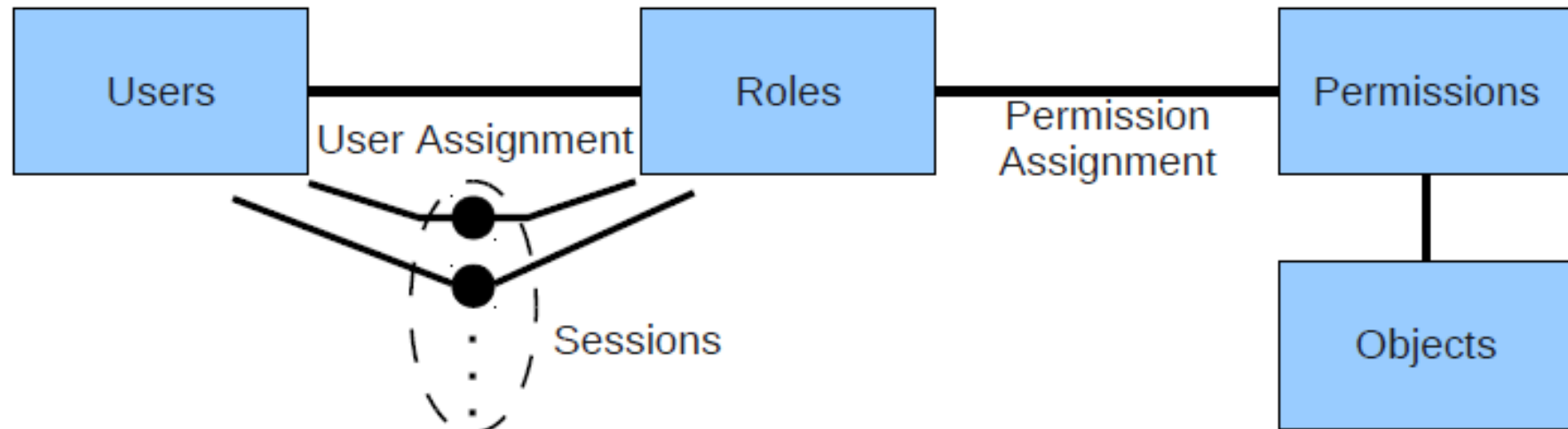
1. Evolve **Simple EM Client** and **GleamViz** to become showcase for **tight integration with Computational Platform**
2. Refine and populate **the catalogue** of epidemic resources: **enrichment, interlinking and semantification of epidemic data**
3. Release second version of the EM.
Re-implemented Web Services (no more Muradora)
New information architecture, new front-end design
New **social network access control**

On the nature of Soc Intelligent Systems

- Who should learn behaviours about individuals from the network?
- No Silver Bullet
- “Classic” Engineering approaches too slow for 21st century pace
- We are now all part of a huge Living Lab
 - How much longer will the fact that your cat sneezed be relevant?...we might have to ask again.
 - Are we still under control?
- We may need more flexible ways to control access to sensitive data..

Classical Approaches

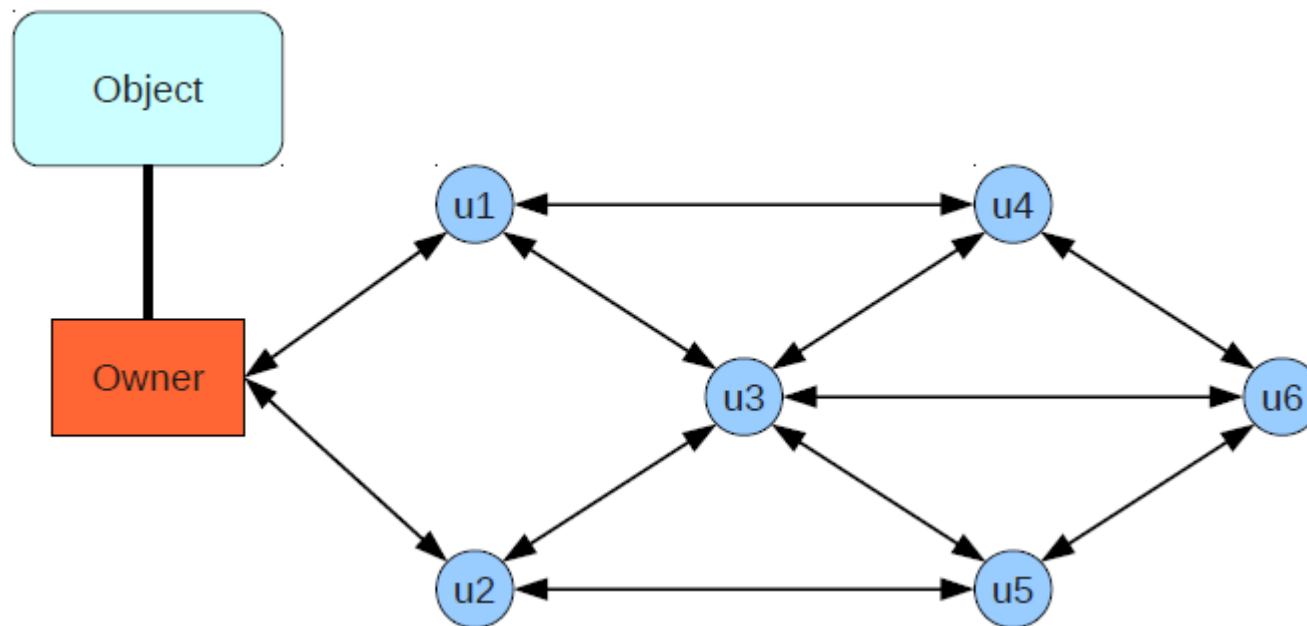
- Role Based Access Control (RBAC):



- Advantages:
 - Roles are intuitive concepts in organizations
 - Users can easily be reassigned from one role to another
- Disadvantages:
 - Central Administration has to manage roles
 - Does not take into account collaborative/social dynamics

Access Control Based on Social Networks

- Objects have owners (or publishers)
- Owners are part of a social network and define access policies based on the network information




EM 2.0 Software Components

- Fedora Commons 3.4 - main features of the repository.
 - Mediator services reimplemented. Webservices provided by FC invoked directly.
- Access control in the platform
 - XACML + LDAP (Tuttle et al. 2004)
 - Shibboleth (identity management).
 - **Access Control Based on Social Networks**
- **Front-end based in the Drupal CMS**
 - Integrated forum



EM 2.0 Mock-up interface


http://v2.epimarketplace.net/mockup


**EM**
Epidemic Marketplace ... a platform for integrating and sharing epidemiological data.


Login


Thu 00.26 AM

// Epidemic Marketplace // Contact // Learn More

**Search**

**Browse**

**Upload**

**Request ?**

Recent Requests (3)
You can also see all the data requests made to date and the most popular data requests.

Recent Uploads (8)
You can check and follow all the recent data uploads made on the Epidemic Marketplace.

Not registered yet? Please Register

Advanced Search
Learn More At la facinci tet adipit loborti


Become Curator

Developers Corner

3000 datasets
Learn More At la facinci tet adipit lobortio consed erit praessis
[See More](#)

9.867 visits
Learn More At la facinci tet adi
[See More](#)

Latests News
Learn More At la facinci tet adipit lobortio consed erit praessis at. Ut eugiam, veliquametue consenis etum quat.
Ostrud erostrud modolent lutpatie ming ercilis am, sum delendip ea at.
Magna con veliquat. Ud eugait lore dolut aut exero
[See More](#)

Website Tour

Guided Tour

Most Visited Documents of the week

Title: At la fancini doko
Author: At la fancini
Date: 2011 / 02 / 13
Description: ijijibmf otjiit dhswyf brb kg morembo fmh dpd [See More](#)

Title: At la fancini doko
Author: At la fancini
Date: 2011 / 02 / 13
Description: ijijibmf otjiit dhswyf brb kg morembo fmh dpd [See More](#)

Title: At la fancini doko
Author: At la fancini
Date: 2011 / 02 / 13
Description: ijijibmf otjiit dhswyf brb kg morembo fmh dpd [See More](#)

Title: At la fancini doko
Author: At la fancini
Date: 2011 / 02 / 13
Description: ijijibmf otjiit dhswyf brb kg morembo fmh dpd [See More](#)

WP3 SWOT Analysis

Strengths

- Epiwork-driven EM
- Standards-based
- Open Source modules
- Supported (until 2012)

Weaknesses

- **Unpopulated EM**
- Looking for the right policies
- What are the incentives?
- Interfaces to WP4 and WP5?

Unchanged !

WP3 SWOT Analysis

Opportunities

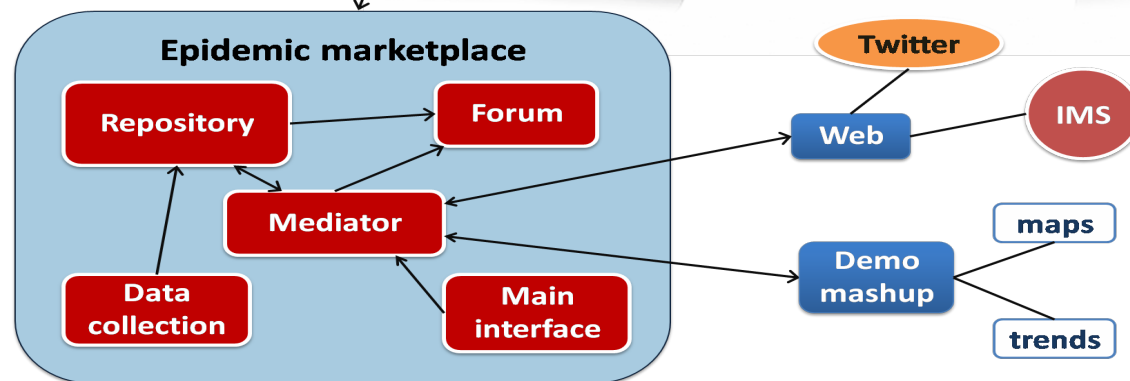
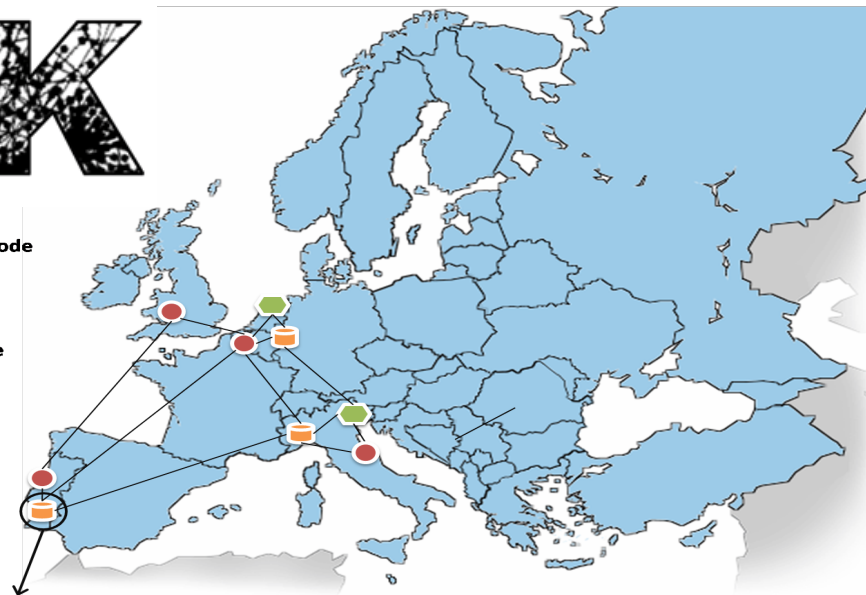
- Epiwork testbed
- Creation of a baseline for epidemic modelling
- Showcase for partners' outputs

Unchanged !

Threats

- Consortium enters “everyone for himself” mode.
- “Somebody will take care of that” attitude
- EM perceived as a very expensive, complex and useless cache

EPIWORK



<http://epimarketplace.net>

